GUIDANCE FOR INDUSTRY

CLINICAL DEVELOPMENT

PROGRAMS FOR DRUGS, DEVICES,

AND BIOLOGICAL PRODUCTS

FOR THE TREATMENT OF

RHEUMATOID ARTHRITIS (RA)

DRAFT GUIDANCE - NOT FOR IMPLEMENTATION

This guidance document is being distributed for comment purposes only.

Draft released for comment on January 6, 1997

Comments and suggestions regarding this draft document should be submitted by February 15, 1997 to Docket No. 96D-0067, Dockets Management Branch (HFA-305), 12420 Parklawn Dr., Rm 1-23, Rockville, MD 20857. Comments and suggestions received after this date may not be acted upon by the Agency until after the document is next revised or updated. For questions regarding this draft document, contact Rose Cunningham at 301-594-5470.

U.S. Department of Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Center for Biologics Evaluation and Research
Center for Devices and Radiological Health

TABLE OF CONTENTS

GUIDANCE FOR INDUSTRY[1]

CLINICAL DEVELOPMENT

PROGRAMS FOR DRUGS, DEVICES,

AND BIOLOGICAL PRODUCTS

INTENDED FOR THE TREATMENT OF

RHEUMATOID ARTHRITIS

This document is intended to assist developers of drugs, biological products, or medical devices intended for the treatment of rheumatoid arthritis (RA) by providing guidance on the types of claims that could be considered for such products and the clinical evaluation programs that could support those claims. Section I addresses types of claims that are available for the treatment of RA and the measures used to support such claims. Section II contains guidance on the timing, design, and conduct of preclinical and clinical trials for RA products. Section III contains guidance specifically pertaining to biological products. Section IV contains guidance pertaining to devices, and Section V contains guidance on special considerations for juvenile RA.

---

[1](When Finished) -- This guidance has been prepared by the Rheumatology Working Group of the Medical Policy Coordinating Committee (MPCC) of the Center for Drug Evaluation and Research (CDER), the Center for Biologics Evaluation and Research (CBER), and the Center for Devices and Radiological Health. Although this guidance does not create or confer any rights for or on any person and does not operate to bind FDA or the industry, it does represent the agency's current thinking on the evaluation of drugs, devices and biological products intended for the treatment of Rheumatoid Arthritis. For additional copies of this guidance contact the Division of Communications Management (formerly the Executive Secretariat Staff), HFD-210, Center for Drug Evaluation and Research, FDA, 5600 Fishers Lane, Rockville, MD 20857 (Phone: 301-594-1012). An electronic version of this guidance is also available via Internet using FTP, Gopher or the World Wide Web (WWW). For FTP, connect to the CDER anonymous FTP server at CDVS2.CDER.FDA.GOV and change to the "guidance" directory. For Gopher connect to the CDER Gopher server at GOPHER.CDER.FDA.GOV and select the "Industry Guidance" menu option. For WWW, connect to the FDA Home Page at WWW.FDA.GOV and go to the CDER section.

I. CLAIMS FOR THE TREATMENT OF RA

Over the past decade, there has been a search for better measures to describe patient outcomes in RA clinical trials. A number of organizations, including the International League Against Rheumatism, the American College of Rheumatology (ACR), and the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group, have attempted to define core groups of measures as well as composite indices describing patient outcomes. As a result of these efforts, several new measures have been described and validated with clinical data. These outcome parameters are now being used in clinical trials during drug development. For this reason, and in the hope that these measures will provide more useful information about patient outcomes, FDA is providing guidance about the use of these new measures in clinical trials that will support label claims.

In addition, many novel agents are under study for the treatment of RA. There is a search for more effective therapeutics that will have a positive impact on the natural history of the disease. The following label claims allow for descriptions of treatment effects of greater benefit than partial mitigation of signs and symptoms.

Although label claims have diverse legal and regulatory ramifications, their central purpose is to inform prescribers and patients about the documented benefits of the product. Because RA is a chronic, symptomatic disease that can result in a variety of adverse outcomes with different chronology, severity, and overall patient impact, various outcomes can be the bases for claims. The claims discussed in this section represent the current views of Agency rheumatologists about achievable and clinically relevant overall outcomes. In addition to the traditional claim of improving signs and symptoms, five further claims are described: improvement in functional capacity/health related quality of life, major clinical response, complete clinical response, remission and prevention of structural damage. More than one claim can be pursued simultaneously. It is anticipated, however, that under most circumstances, any of the additional claims will be approved only if there is adequate evidence to support the signs and symptoms claim.

Given the chronicity of RA, the signs and symptoms claim should be based on trials of at least 3 months duration (trials of biologic agents should be at least 6 months in duration). Claims of improved functional ability/quality of life should be based on trials of at least 6-12 months and all other claims should be demonstrated in trials of at least one year. Some agents, by their nature, need to be evaluated for more than 3 months before a conclusion of effectiveness can be drawn. For example, it is recommended that most efficacy trials for biological drug products be at least six months in duration to assure that the response is durable and not undermined by neutralizing antibodies or other immune regulatory effects. **[FDA is soliciting comments from the Advisory Committee members on trial duration. A number of commenters thought that 6 months is more appropriate for the signs and symptoms claim but there is lack of consensus on how long trials assessing improved QOL should last].**

Given the importance of joint structure in long-term RA management, all trials lasting a year or longer, even if an X-ray claim is not sought, should include a structural assessment (e.g., X-ray, MRI). Trials evaluating claims other than signs and symptoms data should be designed to show superiority, unless active control agents approved for that claim are available.

Claims can be submitted singly or together. Because the persuasiveness of trials showing a difference is, in general, much greater than that of equivalence trials, it is highly desirable for a claim to be convincingly demonstrated in at least one trial showing superiority of the test agent over placebo or active control.

In some instances, a claim of superiority over a specific comparator, rather than a straightforward efficacy claim, will be sought. For example, the desired claim could be for efficacy superior to a specific non-steroidal anti-inflammatory drug (NSAID) for the treatment of signs and symptoms of RA. Substantiation of any claim of superiority over a specific agent should have two adequate and well controlled trials showing superiority. These trials could also be the basis for demonstration of the product's efficacy.

A. Reduction in the Signs and Symptoms of RA

This claim defines symptomatic benefit, or benefit that includes improvement in signs of disease activity as well as symptoms. Ordinarily this claim is established by trials of at least 12 weeks duration (at least 6 months for biologicals). Unless there is a reason to weight symptoms at the last visit more than intermediary symptoms, an analysis which equally weights all time points is appropriate. Acceptable outcome measures that would support claim A include:

1. Validated composite endpoints or indices of signs and symptoms

   These composites can be used to define a categorical endpoint of patient success or failure. For example, the Paulus criteria or the ACR definition of improvement (20% improvement in tender and swollen joint counts and 20% improvement in 3 of the 5 remaining core set measures: patient and physician globals, pain, disability, and an acute phase reactant [1 2]) could be used to assess if a patient responded or not.

   [Illustration: Success for each patient in a six month trial could be defined as meeting the criteria for improvement over baseline in at least four of six observations, and not dropping out because of toxicity.]

2. Well-accepted sets of signs/symptoms measures

For example, the four measures previously recommended in the CDER Guideline for the Clinical Evaluation of Anti-inflammatory and Anti-rheumatic Drugs (1988) [joint counts (pain/tenderness and swelling) and global assessments (physician and patient)] or the ACR core set, may be used as outcome measures. The criteria for success and the methods for statistical analysis should be prospectively defined and agreed upon. For example, in using joint counts and global assessments, ordinarily a statistically significant difference between the control and the treatment group in change from baseline in at least 3 of the 4 measures is used as the criterion for a successful trial.

[Question to Advisory Committee members: what if a sponsor proposed using only one sign or symptom, e.g., joint swelling or patient global assessment? What additional substantiation would be convincing? How many measures are needed to support a plausible claim of relief of signs and symptoms?]

For both the above measures, the 66, 48 or 28 joint count is acceptable.

B. Improvement in Functional Ability/health Related Quality-of-life

This claim should be supported by success in both a validated functional measure for RA and a validated health related quality-of-life measure (either an RA-specific measure or a generic measure shown sensitive to RA), e.g., the health related HAQ and the SF-36. Trials supporting this claim should be at least 6-12 month's duration. **[Question for Advisory Committee members: How long is appropriate?]** An analysis according equal weight to all time points is usually appropriate. Ordinarily, proposals for a functional ability/health-related QOL claim should be for agents that have been shown to also improve signs and symptoms, either in the same or in other trials.

C. Prevention of Structural Damage

Prevention of structural damage is an important goal of RA therapy. Trials evaluating this claim should be at least one year in duration.

The following are examples of outcome measures that could be used to support prevention of structural damage claims.

1. Retardation of X-ray progression, using either the Larsen, the modified Sharp or another index

Radiographic claims should be based on comparisons of films taken at one year (or longer) with those taken at baseline. All randomized patients should have films at both time points, regardless of whether they are continuing on treatment or not. Prespecification of the handling of dropouts is especially important in these trials.

2. Prevention of new X-ray erosions - maintaining an erosion-free state or preventing new erosions

   Trials evaluating this claim would ordinarily use a categorical endpoint to assign a status of progression or nonprogression to each patient.

3. Other measures, e.g., MRI

   Other measures such as MRI could be employed. However, because of the technique's potential for identifying minimal, albeit statistically significant changes, the magnitude of difference that will be regarded as clinically significant should be prospectively agreed upon.

The following set of claims define substantial levels of patient benefit above that required for obtaining a signs and symptoms claim. To obtain one of these new claims, trials should show therapeutic efficacy beyond that of currently approved products.

D. Complete Clinical Response[3]

The claim of complete clinical response defines substantial therapeutic activity with greater benefit to the patient than the mitigation of signs and symptoms of RA. Complete clinical response connotes a benefit requiring ongoing drug therapy while a remission claim (E., below) is the same clinical result maintained off therapy. Both are defined as "remission by ACR criteria" and radiographic arrest (no change by Larsen or modified Sharp methods) over a continuous 6 month period. The 1981 ACR remission criteria require at least 5 of the following: AM stiffness < 15 min, no fatigue, no joint pain by history, no joint tenderness or pain on motion, no swelling of joints or tendon sheaths, ESR<20 (males) or <30 (females). Trials intending to evaluate complete clinical response should be at least one year in duration. Trials evaluating complete clinical response would ordinarily use a categorical endpoint in which a successful patient is defined as above.

E. Remission

The claim of remission defines substantial therapeutic activity with greater benefit to the patient than the mitigation of signs and symptoms of RA. Remission is

defined as "remission by ACR criteria" and radiographic arrest (no change by Larsen or modified Sharp methods) over a continuous 6-month period while off therapy. Remission is not intended to imply cure. Trials intending to evaluate remission should be at least one year in duration.

F. Major Clinical Response

This claim is intended to define a substantial response in patients whose disease cannot remit by the above definition due to existing fixed deformities. The major clinical response claim is defined as a continuous 6-month period of (1) success by a yet to be determined criterion **[Issue for Advisory Committee members: There are several proposals for this. One is an algorithm of the ACR core-set (joint counts, globals, pain, function, and acute phase reactant) defined to "capture" only the best 10% of RA patient database used to derive the ACR 20% measure. An alternative proposal is to allow a prespecified number of joints to be invaluable up front, which would allow merging this category with D and E above. The algorithm will be discussed further at the meeting.]** and (2) radiographic arrest as defined above. This claim is based on statistically significant improvement in response rates above background therapy and, as with the claims of complete clinical response/remission, the trials would be at least one year's duration.

II. CONSIDERATIONS IN RA PRODUCT DEVELOPMENT

The following information on preclinical and early clinical development pertains primarily to pharmaceuticals (drugs and biologicals). The general principles outlined in sections C through F are applicable to devices; however, for information specific to the development of medical devices refer to Section IV in this document. Developers of products that combine therapeutic modalities (e.g., biologics and devices) may request assistance from FDA in designating a lead Center for review of the product. Such requests should be submitted to: Office of the Chief Mediator and Ombudsman (HF-7), Food and Drug Administration, 14-105 Parklawn Building, 5600 Fishers Lane, Rockville, MD 20857-001.

Frequently encountered issues in RA product development include:

(1) Selecting appropriate *in vitro* (animal or human systems) and *in vivo* animal models for screening potentially active agents;

(2) Designing and performing appropriate preclinical safety studies to support the use of a new molecular entity in human volunteers or patients;

(3) Balancing the potential need for therapeutic intervention early in the disease course with the need to avoid exposing patients with mild disease to agents that have toxicities or little record of safety;

(4) Identifying the potential risks associated with combination therapies, particularly those with shared target organ toxicity or potential for pharmacokinetic interactions;

(5) Designing adequate and practical long-term safety monitoring;

(6) Designing trials which definitively show clinical efficacy.

The following sections discuss approaches the above issues.

A. Preclinical Considerations

This section focuses on preclinical issues that are specific to the clinical development of anti-rheumatic therapies. In designing toxicity studies, and the timing of such studies, consultation with the agency is recommended concerning the current recommendations and guidelines that address drugs, devices and biological products. Guidance on preclinical safety testing, addressing the need for and design of toxicokinetic, reproductive toxicity, genotoxicity, and carcinogenicity studies, has been developed by the International Conference on the Harmonization of Technical Requirements for Pharmaceuticals (ICH). These documents are available via the FDA internet home page (http://www.fda.gov/cder or cber). Because biologics can pose unique challenges in animal study design (for example, species-specific binding or immunogenicity of human proteins in animals), there is a specific ICH document under development concerning the safety evaluation of biotechnology-derived pharmaceuticals ("Preclinical Testing of Biotechnology-Derived Pharmaceuticals").

1. Pharmacokinetics

Animal studies of drug absorption, distribution, metabolism and excretion are important during the early IND phase to aid in toxicity study interpretation but need not all be completed prior to Phase 1. Generally, for initial studies in humans, determination of pharmacokinetic (PK) parameters such as area under the curve (AUC), maximum concentration (C_{max}) and half-life ($t_{1/2}$) in animals is sufficient to provide a basis for predicting safe clinical exposure.

Preclinical testing of combinations of drugs (or biologics) to be used in patients with RA is often not feasible before the initial clinical trials since a variety of drugs , including NSAIDs, analgesics, corticosteroids, and

disease modifying anti-rheumatic drugs (DMARDs) are currently used to treat RA patients. To evaluate potential interactions, information on the impact of concomitant therapies on pharmacokinetics may be needed to optimize dosing regimens and to identify potential safety concerns. Metabolic interactions often may be assessed in an *in vitro* system using animal or human liver slices, microsomal preparations, or purified p450 enzymes.

Interactions may also result from the presence of individual- or disease-specific factors, such as rheumatoid factor, which may bind to various monoclonal antibody therapeutics; *in vitro* binding studies which identify patients with high titers may be useful in identifying patients who may exhibit unique pharmacokinetics or patterns of clinical response.

2. Biological activity

The biological activity of a potential anti-rheumatic therapy should be established using multiple preclinical model systems (i.e., *in vitro, in vivo, ex vivo*). *In vitro* screens can utilize cells or tissues derived from animal or human sources and are generally used to select drug candidates that have a desired effect on a molecular target. Such assays can also be used to devise appropriate bioassays for the selected agent. Animals, either healthy, with rheumatic disease (spontaneous or induced) or genetically modified, are subsequently used to determine whether the biological effect can be demonstrated *in vivo*. While the *in vivo* system used should mimic one or more aspects of rheumatoid arthritis or its etiology, it is expected that each animal model will have its limitations.

a. *In vitro:*

Data from *in vitro* studies can be useful in defining the potential mechanism of a drug or biologic and for determining relevance of a particular animal species for *in vivo* assessment of activity or safety. These data are especially useful if a potential surrogate marker can be identified in preclinical studies. For example, if the product is intended to affect the CD_4 receptor on lymphocytes, this receptor may be used as a surrogate marker for both activity and certain toxicities.

Several *in vitro* tests may be utilized depending on the mechanism of action of the drug or biologic. For example, binding assays may be useful for developing receptor antagonists or monoclonal antibodies. *In vitro* functional assays, e.g., platelet and neutrophil aggregation, may be useful tests for identifying inhibitors of inflammatory mediators. Enzymatic

assays (such as *in vitro* or *ex vivo* inhibition of cyclooxygenase, lipoxygenase and phospholipase) may also be useful for determining selectivity for the inhibition of isozymes.

b. *In vivo*

Selection of animal models should be made on the basis of pharmacodynamic responses, similarity of animal disease etiology to clinical disease, and/or to define mechanism-based toxicity. Ideally, products that are targeted for a subset of arthritic patients should be developed in an experimental model(s) that is most relevant to that subset. For example, rats are not sensitive to drugs which inhibit 5-lipoxygenase. Therefore, mouse or rabbit models are more relevant to evaluate the anti-inflammatory activity of leukotriene inhibitors.

The development of rheumatic disease models to allow screening for potential RA drug candidates is encouraged. The following examples are meant only to illustrate some models which are in current use and are not intended to suggest excluding the use of others.

Collagen-induced arthritis (CIA)

Collagen-inducted arthritis is often considered to be a suitable model for studying potential drugs or biologics active in human rheumatoid arthritis because of the involvement of localized major histocompatibility complete class II-restricted T helper cell activation and similar histopathological lesions. Radiographs of joints affected by CIA often show erosive changes similar to those seen in human rheumatoid arthritis. The progressive arthritis often results in RA-like joint deformity and dysfunction. Anti-collagen antibodies, which occur in some patients with RA, develop in the CIA model.

The collagen-induced arthritis model has been useful for identifying immunosuppressants and steroid hormones as well as inhibitors of inflammatory mediators. Since this model can be induced in several animal species it may be especially useful for evaluating drugs that have species-specificity, e.g., leukotriene antagonists and 5-lipoxygenase inhibitors. In addition, while functional tests are not routinely used in this model, incorporation of measures of mobility and joint function may enhance its predictive value.

Naturally occurring arthritis or autoimmune response:

MRL/lpr mice, Biozzi H mice and DBA/1 mice have been used to examine the onset of drug-induced tolerance and immunosuppressant drug effects on autoimmunity. The MRL/lpr mouse model has been useful for evaluating immunosuppressants and hormones.

Rat carrageenin-induced acute model of inflammation:

This model has been useful in assessing anti-inflammatory activity of cyclooxygenase inhibitors. Most of the animal models that involve inflammation in the paw may be used for measuring antiphlogistic action of a drug.

Adjuvant-induced arthritis in rats (AA):

AA in rats has been frequently used for screening non-steroidal anti-inflammatory drugs and inhibitors of inflammatory cytokines as well as antimetabolite-like immunosuppressants.

Streptococcal cell wall-induced arthritis:

This model has been used for developing cytokine inhibitors.

Experimental organ transplant in animals:

This model has been used to identify the activity of immunosuppressants and antimetabolites, particularly those directed at cytolytic cellular immune processes.

Transgenic animal models:

A number of transgenic animal models are being developed for the study of rheumatoid arthritis and may prove useful over the next decade. Some examples include: transgenic mice that carry genes for the env-Px region of the human T cell leukemia virus type I genome, humanTNF, CD4, HLA B-27 etc.

3. Toxicology

Preclinical toxicology studies of a drug or biological product are designed to characterize general and specific toxicity using dosing routes and regimens as similar as possible to the proposed clinical trials with

consideration of the demographics and disease status of the intended patient population. For instance, the prevalence of RA is high in females. Therefore, reproductive toxicity studies should be completed early in clinical development to support the inclusion of women of child bearing age in early phases of clinical trials.

Immunomodulatory or immunosuppressive agents administered to RA patients as monotherapy or in combination raise concerns about the adverse effects of prolonged immunosuppression. For example, malignancies (i.e., lymphomas) are a known risk of long- term, non-selective immunosuppression used for treatment of graft recipients. Investigational drug-related opportunistic infections and mortality related to immunosuppression have occurred in RA patients. Sponsors are encouraged to identify and utilize animal models which may assist in selecting drug candidates that selectively inhibit cells and processes responsible for RA.

Anti-rheumatic drugs are often used in combination in an attempt to improve outcomes and minimize toxicities. However, drug interactions may result in increased toxicity, even at lower than previously evaluated doses of either agent. This concern is especially evident for agents which have long half-lives or non-selective activity, or for drugs which share common target organ toxicity. Preclinical toxicity studies which evaluate the use of combined agents may be helpful in predicting clinical safety hazards. The duration of toxicity dosing of animals is usually linked to patient dosing regimens. Development and validation of *in vitro or* whole animal models is encouraged to address concerns regarding short or long-term patient risk due to immunosuppression.

B. Pharmacokinetic/Pharmacodynamic Strategies

FDA is currently developing specific guidance for the performance of studies to characterize the PK/PD performance of products which should be consulted when it is completed (expected completion 6/97). *In vivo* pharmacokinetic studies are needed to evaluate drug disposition and metabolism, degree of linearity and accumulation, dose proportionality, and, for oral dosage forms, food interactions. [4] Some of these data may be gathered in a single study designed to evaluate a number of parameters. During formulation development, bioequivalence studies linking formulations may be necessary.

Because polypharmacy is common during the treatment of rheumatic disorders, *in vitro* binding studies with blood from patients with active disease should be used as a preliminary screening tool for potential displacement reactions.

For products that may interact with rheumatoid factors, e.g., monoclonal antibodies, the frequency of patients with RF reactive to the antibody, as well as the impact of interactions on the pharmacokinetics of the product, should be evaluated when possible.

C. Considerations in Phase 1 Trials

For general information on clinical development pertaining to most drugs and biological products, see "General Considerations for the Clinical Evaluation of Drugs."[5]

"Phase 1" has two connotations: one refers to the earliest, first-time-into-humans trials, while the other encompasses studies of pharmacokinetics, metabolism, drug interactions, special populations and other clinical pharmacology trials described above. It is expected that both kinds of Phase 1 trials will ordinarily be conducted during the clinical evaluation of therapies for RA. This section is primarily intended to discuss issues related to the first time people are exposed to the drug (including to a particular dose level, or duration of therapy).

1. Settings and investigators

   First-time-into-humans Phase 1 studies should be carried out in institutions with a full range of clinical and laboratory facilities and the patients should be kept under close observation. It is desirable that the trials be under the direction of physicians with experience in early drug development and rheumatology, or that a team of investigators combining experience in rheumatology and clinical pharmacology be employed.

2. Subjects

   First-time-in-humans drug trials are frequently conducted in healthy volunteers. Such studies are predicated upon the ability to perform, and to interpret the results of, preclinical animal tests. If the preclinical testing does not reveal potential mutagenic, immune system or potentially serious effects at or near the expected therapeutic range, testing in volunteers is initiated. However, for biological and drug products that have potentially serious toxicities, it may be appropriate for initial testing to be performed in patients with some potential to benefit. This has created challenges in selecting an appropriate initial patient population.

   For drugs and biologics that have been tested in relevant preclinical toxicity evaluations and have been found relatively safe, without the potential for mutagenic, immune system or other serious effects at the proposed doses,

trials may be initiated in healthy volunteers. If however, significant effects have been demonstrated or might be possible, selection of an appropriate patient population is necessary. It is recommended that patients meet the ACR criteria for both diagnosis and activity of RA and be without other serious medical conditions. Patients with minimal disease are sometimes not appropriate for the same reasons that the testing is not initiated in healthy volunteers. Patients with devastating RA may also not be the best starting population because of the medical complications of their disease. In addition, they may be less likely to respond to therapy.

There is ongoing epidemiologic work on identifying markers of increased risk in RA: these could be useful for identifying patients with poor prognoses who might be considered for very aggressive treatments (e.g., immunoablative therapies followed by stem cell transplants) of potential high toxicity. Application of epidemiologic studies may allow a very aggressive treatment to be restricted to a subset of RA patients who have a demonstrated shortened lifespan due to their disease, e.g., subjects with greater than 30 affected joints or a score on the HAQ with fewer than 75% of questions answered "with ease."

In any case it is particularly important that informed consent be complete and that some provisions be made to assess that patients understand what they are consenting to. If the potential exists for disease exacerbation, this should be part of the informed consent.

When the characteristics of the agent suggest that it may potentially have long-term gonadal effects, it is desirable that men and women not wishing to parent children be chosen for Phase 1 studies.

3. Trial design

Ordinarily, initial Phase 1 studies are sequential dose escalation trials, in which safety and tolerance at a specific dose is established before exposing additional subjects to a higher dose. A single dose is almost always tested first, followed by repeated dose studies; however, this design is influenced by the type of agent used. Although escalating the dosage to a clearly determined maximum-tolerated-dose (MTD) will aid future trial design, in some instances it is not medically prudent to try to fully characterize the MTD. Additionally, for some products, an MTD may not be definable.

The starting drug dose chosen is often a "no adverse effect" dose (determined by interspecies mg/m^2/day dose conversion from animal to human). For biologicals, the initial dose chosen is often one thought to

have no adverse biologic effect. Conservative dose escalations (e.g., half log or less), are usually recommended.

4. Concomitant therapy

Use of low-dose corticosteroids (up to 10 mg prednisone equivalent daily), and NSAIDS may ordinarily be continued in Phase 1 trials. Concomitant therapy with methotrexate and similar agents should be avoided in initial phase 1 trials of all novel antirheumatic drugs, biologics and devices because of the difficulty of differentiating the toxicity of the novel agent from that of the co-administered product.

Physicians now prescribe methotrexate and similar agents earlier in the course of rheumatoid arthritis. Recruiting adequate numbers of patients not taking these agents may be difficult. Approaches which may allow the use of methotrexate and similar agents in later Phase 1 trials include: (a) obtaining reassuring evidence of lack of toxicity from relevant animal models in which co-administration occurred; and (b) starting at doses significantly lower than the "no adverse effect level" of the single agent as determined by preclinical studies. Such proposals should be discussed in the planning stages with Agency staff.

5. Observations

a. Safety

The standard batteries of safety observations have been described elsewhere. However, additional types of safety observations may be necessary, e.g., tests of effects on cellular and humoral immune function or host defenses. For products with the potential for effects lasting long after administration, or for delayed toxicity, appropriate follow-up should be designed. For example, Phase 1 studies of agents used to deplete or modify the function of T-cell subsets should be designed to carefully assess both the short and long-term effects on number and functional status (e.g., DTH responses) of cell populations and other pertinent pharmacodynamic assays during therapy and during follow-up.

It is desirable to incorporate individual patient adverse event stopping/withdrawal “rules” into protocol designs. In addition, incorporation of stopping or modification rules for adverse events into trial designs is often advisable. For example, dose escalation rules should be clearly defined in dose-finding studies, with

provisions for enrollment of additional patients at a given dose if possible significant adverse events are observed at that dose.

b. Efficacy

Developing an understanding of the agent's therapeutic potential in early trials is highly desirable for efficient product development. This may be attempted in Phase 1, but can only be achieved by performing controlled trials. RA responses in open trials are of questionable value in indicating efficacy. Consideration should be given to the more modest goal of determining whether the pharmacological effect predicted from the preclinical development is present (proof-of-concept).

D. Considerations in Phase 2 trials

During Phase 2, larger, often longer-term trials are employed to better define the dose- and exposure-related activity and toxicity of the agent. Enough information should be generated to ensure that the Phase 3 trials can be conducted safely and with a high probability of success. In addition, Phase 2 trials should solidify a total drug development strategy, to ensure that, after the Phase 3 safety/efficacy trials are done, all of the information needed for registration will have been gathered, including an appropriate safety database, clinical pharmacology, dose response data, the exploration in special populations (e.g., renal failure, hepatic failure), and drug interaction information with agents expected to be co-administered.

There is nothing to preclude conducting additional "Phase 1" clinical pharmacology studies and Phase 2 trials while the Phase 3 development is ongoing.

The following issues are important for Phase 2 trials in RA:

1. Dose finding

This is a central challenge of Phase 2 development. Once a reasonably safe range of doses has been established, randomized, parallel arm dose-comparison trials are ordinarily recommended. The use of a placebo arm is desirable for several reasons. First, if no difference is found among doses, there is usually no other way to determine whether all doses were equally effective or equally ineffective. Second, if a dose-response trend is found, the placebo arm may indicate the possible magnitude of the observed effect. If use of a placebo is not possible, designs should include wide dose ranges (durations, repetitions, etc.). Active-controlled designs that

specify an arm with a well-characterized, known therapy can also be very useful.

Signs and symptoms measures may be used for dose finding studies, i.e., it is not contemplated that separate dose-finding be done for the longer-term endpoints.

For agents that are thought to have prompt action and rapid offset of effect, alternative designs, including cross-overs and titration designs, may be useful, although historically this has not been the case. Trials of two or more doses which permit liberal titrating per the patients' responses are unlikely to clearly demonstrate a dose response, because these titration designs result in a blurring of any real dose distinction that may exist.

The desirability of identifying a range of doses with acceptable toxicity and reasonable activity, for study in Phase 3, cannot be stressed enough.

2. Safety

Every RA investigational therapy raises safety concerns. Whenever there is a potential for significant, long-lasting or delayed-onset toxicities, it is desirable to design the Phase 2 studies to provide a group of patients with longer-term follow-up preceding the larger Phase 3 studies. Provisions for long-term follow-up can be helpful in addressing issues prior to approval/registration (e.g., issues relating to the potential for immunosuppression, opportunistic infections, neoplasia, and induction of autoimmune disease).

It is desirable to develop a standardized toxicity grading scale for use in all trials of a product, based on the known and suspected toxicities of the product, or of the drug class. This scale may be developed in early Phase 2. This may improve consistency of adverse event reporting, and allow more accurate comparisons among trials.

3. Additional development aspects

a. Concomitant therapy

Before starting Phase 3 trials, an evaluation of the test product's interaction with the other agents likely to be used by the target population should be performed. Initial information can be established based on metabolic pathways, studies of in vitro systems, animal or human pharmacology studies, or drug

interaction studies. This type of information is helpful in directing areas in need of clinical evaluation. When products are intended to be tested as combination therapy with the investigational agent, substantial information on interactions and safety of co-administration should be developed in Phase 2.

b. Gender effects

Most RA trials have predominantly female enrollment. Sponsors should evaluate whether the observed safety and efficacy findings are restricted to women or can be also extrapolated to male subjects. This may be accomplished by subset analyses from trials, PK data, or other information.[6]

E. Efficacy Trial Considerations

The overall goal of Phase 3 work is to demonstrate the efficacy of the product in convincing controlled trials, and to accrue a sufficient safety database. Efficacy trial protocols should contain an analytical plan that precisely identifies the primary comparison(s) to be made, the criteria for success of the trial, and the statistical tests that will be used. These should be linked to the labeling claim that would be supported by the trial. Any additional planned, ongoing, or completed trials that are also intended to support the claim should be identified.

1. Global considerations

a. Patient selection

1) Activity of disease: Unless some other specific subgroup is targeted, patients enrolled in efficacy trials should at a minimum have disease definition and disease activity as defined by ACR criteria. Consultation with the Agency on the generalizability of claims derived from trials with significant limitations on entry criteria is recommended.

To enhance the power of the trial, strategies to improve the chances of a response to therapy are often employed. Some designs incorporate an attempt to select active patients by withdrawing background treatment and allowing patients to "flare". Only individuals with sufficiently high scores are enrolled. The relevance of this type of observed flare is questionable and its ability to predict active disease has not been established. Many patients randomized to placebo in

such studies exhibit the characteristic response of rapidly returning almost to baseline without further treatment. In addition, when patients undergo blinded withdrawal from therapy within these trials, similar dramatic flares are not observed. This raises the question of whether there is an expectation bias on the part of patients, who have been told about the flare procedure, and ascertainment bias on the part of investigators, who wish to have patients meet the entry criteria and enroll in the study. These uncertainties and instabilities around the outcome measures used in such trials should be kept in mind when employing these designs.

A proportionately smaller, but nevertheless noticeable and prompt "regression to the mean" is noted in the joint scores of patients required to have a certain minimum value for trial entry in trials not employing a "flare" strategy. This means that patients, on the whole, will not actually be as active as anticipated when the entry criteria are set. The mechanisms are similar to the above example.

2) Subgrouping patients by disease markers: RA is likely composed of a number of more or less distinct diseases delineated by a common genetic background, corresponding clinical manifestations, similar serologies, and responses to therapy and prognoses. The study of RA may be enhanced by using more homogeneous groups defined by markers with clear prognostic significance. Novel epidemiologic and molecular genetic approaches may lead to identification of even more subgroups. However, prospective studies are first needed to confirm the clinical usefulness of new purported prognostic factors. Where existing data do support markers as prognostic indicators (risk factors), the presence of rheumatoid factor, erosive or vasculitic disease, and DR4 homozygosity, should be taken into consideration in the design of trials. Although in some cases such studies may limit generalizability and impact labeling of the final product it is also possible that such targeting may improve the risk/benefit profile.

b. Concomitant antirheumatic therapy

Studies in RA patients, except in those with very mild disease, are carried out in the presence of concurrent active therapies, including

steroids, NSAIDS, hydroxy chloroquine, etc. This concurrent therapy creates numerous challenges in patient selection, toxicity monitoring and clinical trial design. For example, since methotrexate therapy is used to treat many RA patients, new agents will be used in combination with methotrexate in clinical practice, unless a contraindication exists. Therefore, unless a prohibition on concurrent methotrexate is supportable, data regarding use of the investigational agent in combination with methotrexate is necessary to evaluate the potential for immunosuppression from combination therapy. Other agents may need to be similarly evaluated.

In addition, patients can be categorized according to their responses to standard therapy. Varying trial designs may be required to assess the response of different subgroups to an investigational therapy. For example, with respect to methotrexate use, the RA population can be divided into five groups: (1) methotrexate non-candidates - disease too mild or too early for methotrexate; (2) methotrexate candidates - disease sufficiently (or will become sufficiently) active to justify methotrexate; (3) methotrexate successes - disease reduced to negligible amounts; (4) methotrexate failures - clear drug failures, for inefficacy or tolerability, and (5) methotrexate "partial responders" - with considerable residual disease despite methotrexate. Each of these groups might be considered separately for candidacy for an investigational agent, and with respect to an appropriate trial design. If only a subpopulation of RA patients (e.g. methotrexate non-responders) is studied in a particular trial, the results would ordinarily reflect efficacy only in that group. Any planned subpopulations should be clinically distinguishable. Sponsors should consult Agency personnel when planning a clinical development program contemplating an RA claim that is limited to a subpopulation of the disease.

c. Other Concomitant Therapies

Most patients with RA are taking concomitant medications. Use of medicines unlikely to influence treatment outcomes (e.g., antihypertensives) should simply be recorded, although investigators should be alert for possible drug interactions. The following approaches may be considered in dealing with arthritis medications or analgesics. Obtaining information during clinical

development on co-administration of the test medication and expected concomitant medications is desirable.

1) Prohibit their use. This strategy may result in noncompliance or an increased number of dropouts.

2) Incorporate protocol-specified use, with monitoring. With this strategy, additional analgesic use (and possible other arthritis medications) may be used according to protocol specified criteria. In addition, for long duration studies, protocols should address whether intra-articular steroids are permitted and, if so, for how long assessments of the involved joint are excluded from analysis, and the manner in which "stress" doses of corticosteroids for surgery, etc., are to be handled and how soon after such doses protocol assessments would be allowed.

3) Design analgesic use, or its quantitative consumption, as (part of) an efficacy endpoint.

4) Define use of more arthritis treatments as (part of) an efficacy endpoint, or as (part of) a definition of treatment failure.

d. Stratification

Randomization is intended to balance confounders; however, in any specific trial, especially a small one, randomization may fail to achieve balance. It may be advisable to stratify known (or highly suspected) major risk factors to ensure their balance across arms. Any factor whose influence on the outcome is suspected to be as strong as the treatment's influence should be considered for stratification (e.g., erosive disease, presence of rheumatoid factor). An often overlooked risk factor is the patient's past therapeutic history. (See statistical section for further discussion)

e. Blinding

Because most RA outcome measures have a high degree of subjectivity, full patient and assessor blinding are usually needed for a credible inference. Designs may have compromised blinding if there is not an approximate parallelism in time to onset, nature of response, and toxicity profile. Trials should have parallel dosing in

both arms so that a drug requiring frequent dose manipulations does not threaten the blind. If "arm specific" treatment adjustments are necessary, e.g., per monitored drug levels, these can be done by an unblinded (and sequestered) third party, in order to maintain patient and assessor blinding. Similarly, if the blind is likely to be compromised by infusion related events or other features of the treatment protocol, critical treatment endpoints such as joint counts should be assessed by an independent party with no knowledge of the subject's history.

f. Effects of dropouts and noncompliance.

It is important that trials be designed to minimize dropouts and the attendant information loss. Traditionally, recommended RA trial designs have focused on eliminating sources of variability, for example, extra pain medications, intra-articular injections, etc. Often, these treatments constituted a major protocol violation, requiring that the patient be dropped from the study. There is a trade-off between patient retention and tolerance of variability in RA trial design. Protocols demanding rigid adherence may yield uninterpretable results because of dropouts and noncompliance emanating from patient and investigator intolerance of the requirements. On the other hand, protocols permitting any kind of additional intervention may likewise be so confounded as to defy interpretation.

The following strategies may help minimize loss of information:

1) Use screening or run-in periods so that patients are randomized to treatment groups only after their eligibility and commitment is confirmed.

2) Thoroughly train investigators and study personnel to minimize inappropriate enrollments, protocol violations, and other deviations that would decrease the ability to assess trial outcomes.

3) Include dropout in the definition of the endpoint, as in a time to defined treatment failure, or a defined by-patient success or failure.

One example of this approach would be to use a protocol defined response rate as the primary endpoint. Dropouts,

and patients not dropping out, but having minimal or no response to therapy, are classified as nonresponders. With this type of endpoint, the criteria for classification as a nonresponder need to be clearly and prospectively defined. **[Issue for Advisory Committee members: Is this approach appropriate?]** [Illustration: In a study of 6 months duration the primary endpoint could be a comparison of the proportion of patients with an ACR 20 response at six months. Such a protocol might specify that if no more than 15% improvement compared to baseline were seen on two consecutive study visits after two months on protocol, the subject would be declared a nonresponder. Nonresponders could be removed from study drug, and changed to an alternative treatment if desired by physician and patient, but would continue to be followed until the end of the study.]

4) Make provisions for following patients who have stopped experimental treatment. Options include allowing a protocol specified crossover to a standard therapy, for patients meeting predefined criteria for treatment failure.

5) Allow more flexibility in treatment options during the study. Some designs that have been used include allowing dose adjustment of the comparator arm (assessor and patient blinded); allowing add-on therapy for patients meeting predefined criteria for inadequate response, and allowing a limited number of joint injections, with elimination of that joint from assessment.

2. Trial designs in RA

Clinical trials in RA can be designed to test a difference - demonstrating that the investigational product is superior to control (placebo, lower test dose, another active agent), or they can be designed to test an equivalence claim - demonstrating no difference in efficacy from active control. Placebo-, dose-, concentration or active-controlled designs are acceptable. It is desirable that at least one study show an unequivocal treatment effect, i.e., the test drug has better efficacy than a randomized control arm, whether the control arm is a lower dose of the agent, an "active" control, or a placebo.

a. Superiority trials

The standard two arm, investigational agent versus placebo design has been the most common RA design and is the most straightforward. The details of trial design will depend on the population tested. Patients with mildly active RA taking only NSAIDS, who have never been treated with an additional class of therapy, may be enrolled in a placebo-controlled trial with continuation of NSAID background therapy; however, patients doing poorly on NSAIDs alone are usually not appropriate candidates for placebo controlled trials. The same considerations apply to patients who are partial responders to, or who have failed, various other treatments.

Alternative versions of the two arm difference design are a standard dose response study, and a superior to active control hypothesis. These designs may accommodate the need to provide active treatment to patient groups where randomization to placebo is not feasible.

b. Equivalence trials

Equivalence trials are designed to support a claim of effectiveness by showing that the investigational drug is most likely as effective as an active control. The criteria for determining equivalence should be prospectively stated and should be based on achieving 95% confidence that the real difference is smaller than a predetermined amount. Standard confidence limit statistical techniques should be used. Achieving similar point estimates of the efficacy of the two agents is ***not*** a demonstration of equivalence. Equivalence trials usually require more patients to achieve adequate power than difference trials.

A major problem in equivalency trials is assuring that both treatments were equally effective rather than equally ineffective. Approved agents for RA have fairly small effects and frequently fail to show efficacy when tested against a placebo. Comparative trials intended to show "equivalence" to such treatments, when not anchored by a placebo control group, may lack credibility. It is desirable in equivalence designs to select highly effective comparative agents used in the optimum dose and patient population. If possible, use of a third (placebo or lower dose) arm, so that a treatment difference can be shown, is a desirable strategy in equivalence trials. This arm would not necessarily have as many

patients or as long a duration as the active comparators. It is important to design both efficacy and safety measures in a manner that is not biased against the control to ensure a "fair comparison."

Trial conduct that adds to the inherent variability in the outcomes may obscure differences and thus lead to a false conclusion of equivalence. This is the opposite of a difference design, where sources of variability work against trial success. For this reason, minimizing dropouts, patient non-compliance, and missing data is essential to the credibility of the study.

[Example of a statistical equivalence test: As an example of these design decisions, consider the setting where response rates to methotrexate (in methotrexate candidates) with measures such as the ACR 20% are estimated to be on the order of 50%. In this setting, new agents studied in equivalence designs with the methotrexate control might, for example, be expected to show a responder rate around 50%, with a 95% confidence interval or window in the range of up to ± 20%. In other words, if the agent shows the lower bound of the response rate within 20% of the active control response rate result, and if both the test and methotrexate statistically exceed the response rate for a negative control arm, equivalence would be declared. 80% power calculations to determine sample sizes, given the null hypothesis of not more than a 20% difference of two agents assumed equivalent, yields a figure in the range of 125 patients per arm.]

As noted above, requirements for patient number and/or trial duration are usually more demanding for equivalence trials compared with difference trials. Proposals for equivalence trials will be considered by the Agency on a case-by-case basis, depending on the particular agent of interest, the positive control used, the outcomes measured, and the patients enrolled.

c. Trial designs novel to the study of RA

The following designs have not been traditionally used in the study of new RA treatments, but may be considered in certain circumstances.

1) Withdrawal designs. The withdrawal design -- in which patients in both arms of a study are treated with the investigational agent, which is then blindly withdrawn from

one arm, after which patient outcomes are compared -- is sometimes used to assess efficacy. Demonstration of statistically significant worsening in patients taken off the investigational drug demonstrates effectiveness. Natural endpoints for withdrawal designs are "time to (predefined) worsening" using standard "time-to-occurrence" statistical tests, or a simple comparison of proportion of outcomes in the two arms. Withdrawal studies may be performed with both arms on background therapy.

2) Induction designs. **[Issue for Advisory Committee members: FDA would like advice on the evaluation of short-term administration of agents that are intended to have longer term results--hence the term "induction."]**

3. Analytical Issues

a. Handling Dropouts.

Historically, inferences from RA trials have suffered from diminished reliability because of information loss due to dropouts. Dropouts probably never occur randomly, and rarely occur fully independent of the treatment being tested, so there is always the possibility that dropouts introduce a bias. This problem is common in many randomized trials. There have been methods proposed for analyzing the effects of dropouts, but none is fully adequate. An approach with the potential to deal with this problem is to follow all patients, including dropouts, to the planned trial endpoint (even if post dropout information is confounded by new therapy).

This problem is not solved by using the "intent-to-treat" (i.e., all randomized patients included) analysis with an imputation by "last observation carried forward" (hereafter called ITT/LOCF), nor by showing that ITT/LOCF and PP/OC (per protocol completers/observed cases only) analyses concur.

Thus, the effects of dropouts should be addressed in all trial analyses to demonstrate that the conclusion is robust. This may be accomplished by showing the result holds despite application of the "worse case rule" - assign all post-dropout scores for placebo patients the best score, and for all for the drug patients the worst score.

b. Comparison to baseline outcome measures

A phenomenon frequently observed in RA, as well as other conditions, is that patients who stay in trials do better than those who drop out: "Responders do better than non-responders." This is true for both placebo groups and active treatment groups. If observations of the disease were made exclusively from clinical trials, one might conclude that the natural history of the disease is inexorable improvement. This phenomenon is attributable to preferential dropout of worsening patients (a phenomenon not adequately compensated for in LOCF analysis) as well as "regression to the mean." The problem is exacerbated in flare designs, where all patients have major improvement regardless of treatment status. This fact makes comparison-to-baseline outcome measures very difficult to assess, since, very often, much of the improvement noted has no relationship to a treatment effect. For these reasons, active controlled trials not incorporating a placebo arm, and using comparisons to baseline, may be extremely difficult to interpret, especially if a flare design is employed.

4. Statistical Considerations in Efficacy Trial Design

It is advisable to discuss the design and analysis with the FDA review team prior to embarking on a study. In addition, FDA's Guideline for Format and Content of the Clinical and Statistical Sections of New Drug Applications contains useful information.

a. Randomization/Stratification

The purpose of randomization is to allocate patients to treatment groups to assure that unbiased estimates of differential treatment effects exist, since it is not possible to predict all influential factors.

In some clinical trials, there are known factors that are at least as influential in controlling the observed severity of disease as the drugs being studied. Stratification may be used to avoid relying on randomization properties to balance patient assignment for these factors. Stratification is implemented by constraining simple randomization to balance the assignment of patients to treatment groups for the chosen stratification factors.

Every Phase 2 and Phase 3 study protocol should contain a randomization section. All constraints imposed on the

randomization should be explicitly identified. It can then be inferred, when a stratification factor or sample size allocation constraint is not mentioned in a protocol, that there exists no corresponding randomization constraint. This applies to whether patients are blocked to balance treatment assignment for time of patient entry into study and to the more obvious stratifications on center and baseline.

Because stratification implies constraints on randomization, studies that have been stratified for certain factor(s) should account for these factors in the statistical section of the study protocol. The protocol analysis should be implemented for each study.

There are also statistical procedures to address bias in treatment group comparisons by adjusting for imbalances in pre-specified factors (covariates).

It is not required that randomization be stratified; however, failure to stratify can be unwise. In all clinical trials, practical judgment is required in deciding when to stratify. There are reasons to choose stratification and reasons to choose statistical adjustments.

1) The advantages of stratification are, first, that it is better to avoid possibly major statistical adjustments of differential treatment effects. Stratification would essentially eliminate the effect of such adjustments before analysis began. Second, although stratification and statistical adjustment procedures are both designed to remove bias in estimated treatment effects, stratification is more powerful. This is because stratification leads to smaller variances of estimated treatment effects than does statistical adjustment without stratification. Finally, the inclusion of stratification factors into a statistical analysis model should result in increased power to detect effectiveness.

2) Stratification becomes increasingly clumsy as the number of strata increases, and consequently, the available number of randomizable patients per cell decreases. It is logistically simpler not to stratify, relying on statistical methods to adjust for the minor imbalances usually resulting from failure to stratify.

The best approach may be to combine stratification, applied to a limited number of the most influential prognostic factors, with statistical modeling. Statistical modeling would account for stratification and would be used to adjust for the effects of a parsimonious number of the most important remaining factors.

b. The Identification of Primary Efficacy Variables

Each Phase 2 or Phase 3 study protocol should identify the primary and secondary efficacy variables. Primary efficacy variables are critical to the identification of the effectiveness of the product. It is for the primary efficacy variables that statistically significant results are expected to confirm the superiority or the clinical equivalence of a product. Secondary efficacy variables are those which support the validity of the primary variables but which are not critical in deciding if this product is effective. It is helpful, but not necessary, that statistical evidence of efficacy be shown for secondary efficacy variables.

c. Prespecification of Statistical Analysis

Statistical analysis of primary clinical endpoints is part of the process for obtaining consistent and convincing evidence of product efficacy. These statistical analyses should not be data driven. In part, this is implemented by identifying, in each study protocol, before data are available for analysis, a sufficient description of the statistical analyses of these primary efficacy variables so that an independent statistician could perform the protocol analyses. This description of the statistical analyses should include but not necessarily be limited to specifying (1) what constitutes the minimal statistical results needed to demonstrate a successful outcome, (2) whether statistical tests of hypothesis or confidence intervals will be 1- or 2-sided, (3) what level of significance is to be used, (4) how missing values and dropouts are to be handled, (5) the mathematical expression of the statistical model used, and (6) the planned multiple treatment comparison method.

d. Multiple Endpoints

1) Many RA studies use multiple endpoints to assess primary evidence of effectiveness. For example, for the four measures recommended in FDA's previous guideline, trial results were considered to support a conclusion of effectiveness when statistical evidence of efficacy was shown for at least 3 of the 4 measures: physician global assessment, patient global assessment, swollen joint count, and painful joint count.

2) Multivariate statistical methods are also available for analyzing the set of primary efficacy variables.

3) Efficacy variables can be combined within patients (composite endpoint). Such a fixed combination of efficacy measures should be well defined in the study protocol. Composite efficacy variables have the advantage of avoiding several statistical and inferential difficulties associated with multiple endpoints.

e. Dropouts

Dropouts are patients who, after a certain period of time in a trial, fail to provide clinical efficacy data scheduled by protocol to be collected. Frequently, dropouts occur for reasons related to taking the assigned test drug (adverse effects or lack of efficacy). Since dropouts do not usually occur randomly, the remaining patients constitute a biased sub-sample of the patients originally randomized.

Methods used to handle dropouts, such as the "LOCF" and "completers" analyses are not fully satisfactory even though they have often served as the basis for determining that adequate statistical evidence of efficacy has been provided. The LOCF method generally does not preserve the size of the test, either for the comparison of final observations or for the comparison of rates of change. Alternative methods include growth curve analysis and random effects regression. These are also susceptible to informative censoring--that is, dropping out depends on the value of the response. It is often useful to show that the results hold for a variety of analyses--i.e., they are robust.

f. Trials with Several Treatment groups/Multiple Comparisons

In clinical trials involving more than two treatment groups, a statistical multiple comparison procedure controlling the experiment-wise error rate to 5% or less should be applied. In essence, there should be overall statistical evidence of a treatment main effect before attempting to make specific drug comparisons relevant to proposed drug labeling.

g. Interim Analyses

Interim analyses are those which, for any purpose, are performed on partially accumulated clinical trial efficacy data. The study protocol should state whether such interim analyses are planned or not planned. Should interim analyses be planned, that plan and its implementation should be described in the protocol. The protocol should identify the scheduling of these analyses, the method to be applied for adjusting significance levels, and the corresponding time sequence of significance levels at which statistically significant results will be claimed.

While an interim analysis may not be thought to affect the subsequent collection of efficacy data, interim analyses carry an additional risk that the blinding or conduct of a study may have been compromised. Because multiple tests (including interim analyses) alter the true significance level, methods have been developed to compensate for this phenomenon. These statistical methods cannot compensate for any unblinding and bias that may result from gathering the information needed to perform an interim analysis.

h. Sample Size

Failure to recruit an adequate number of patients is a major reason why an effective drug product may fail to meet established statistical criteria for efficacy, independently of whether the purpose was to show superiority or comparability of treatment effect. The method of determining the sample size should be stipulated in sufficient detail to permit independent verification of the computation. This should include identifying the efficacy variable the sample size determination is based upon, the magnitude of the

clinical difference to be detected, the power, the significance level, and the sidedness of the statistical procedure(s) described in the analysis plan. Furthermore, the size of the clinical difference chosen should be justified and the choice of the efficacy variable used to determine sample size should be discussed briefly.

i. Trials to show Clinical Equivalence

The words "clinical equivalence" are used in a much more narrow sense than these words might imply to the casual reader. First, there is often no intent of showing equivalence of two or more drugs across the broad spectrum of pharmacologic effect. Rather, focus is on showing no clinically relevant differences for one or possibly more variables which are to be clearly identified in advance. The concept of equivalence is two-sided in that if, for any outcome measure, one drug is sufficiently different from another drug, then these drugs are no longer deemed equivalent in that variable.

To show equivalence, the variables serving to measure these effects of interest should be defined in the protocol. For each efficacy variable for which clinical equivalence of effect is sought the magnitude of a difference deemed to be inconsequential should be identified. The clinical data should then show, with 95% confidence, that this pre-defined difference is not exceeded.

Inference based on trials to show equivalence is inherently less convincing than inference based on trials to show the existence of a difference. Often clinical trials do not detect treatment differences which are known to exist. In such cases, statistical methods may then seemingly provide evidence of equivalent effect, e.g., to placebo.

In cases where a per patient success rate can be established, equivalency may be demonstrated if the two sided 95% confidence interval of the test group does not exceed ±10% of the control group rate (±15% of a control rate of 85% or lower, or ±20% of a control rate of 80% or lower). [**Issue for Advisory Committee members: FDA seeks your advice on these intervals. Is potential loss of 20% of the active control effect acceptable? Is**

the interval too tight? Is the reference to scales understandable?]

In cases where individual scales are used, 95% confidence intervals that are contained within a 10% range (around the control value) of the total used portion of the scale are generally recommended.

j. The Role of Statistical Significance

Drugs are approved on a weighing of risks and benefits. Rejection of a null hypothesis of no drug effect is evidence that a drug effect does exist. This does not necessarily imply that the effect thus detected is adequate. The magnitude of difference in drug effect that is clinically meaningful should be addressed in the protocol and discussed in advance with FDA representatives.

k. Types of efficacy endpoints

The goal of the statistical analysis of the endpoint is to demonstrate if the product shows convincing evidence of efficacy. Studies of RA generally involve measurements taken at several times, and statistical methods appropriate to this design need be employed. The primary efficacy variables should be specified in the protocol for the study and the proposed analysis should be outlined. In the analytical plan, the method of determining the sample size should be stipulated in sufficient detail to permit verification of the computation. There are several options for endpoints available:

1) The response may be a binary variable indicating improvement from baseline. The analysis here has a straightforward interpretation if all patients are included to completion. If some patients have only partial follow up, it may be unclear how they should be scored unless the procedure is specified and justified in the protocol.

2) The response may be an ordered categorical one (e.g., much worse, worse, no change, better, much better). Such responses are usually analyzed using ranks (accounting for ties), leading to a Wilcoxon rank-sum test. The response is measured at the specified ending time of the patient regimen. If patients fail to complete the regimen, there is no clear way to impute the subsequent time point data.

3) The response may be a continuous variable (e.g., time to an event, the tender joint count), and the difference between final and baseline used ("change score"). This widely used method has the advantage of measuring a clinically recognized difference, but it does not account for time. By dividing the change score by the time interval, a rate of change per unit time is obtained, which allows inclusion of data from all patients whether they complete the study or not. Similarly, one could determine the best fit slope for each patient's measurements.

l. Appropriateness of the statistical methodology

The appropriateness of the statistical model should be assessed, including checking for outliers and determining if distributional assumptions (usually normality) are met and if common variance assumptions hold homoscedasticity.

m. Site effects

If the patients have been stratified and randomized by site, the analysis should include a site effect. There may be a site by treatment interaction reflecting the degree to which the treatment varies across sites. This is often notable when there is a great variation in enrolled patients across sites. Site by treatment interaction should be explored.

F. Safety Analysis

The approach to evaluating adverse reaction data and laboratory values has traditionally differed from that used to evaluate efficacy. The purpose of safety evaluations is usually not to test a specific hypothesis, but rather to examine the pattern of effects and to detect unusual or delayed events. Analyses using cumulative occurrences, scatter-plots of laboratory values (baseline versus on-therapy), or general regression techniques may be helpful. The safety profile should address to what extent adverse events (drug reactions or lab values) depend on duration of drug exposure, dose level, coexisting medical conditions, or possible drug interactions. Incidence rates should be calculated using

denominators that reflect the period of drug exposure for the population at risk. Cumulative incidences (hazard rates, instant probabilities) better represent the temporal pattern of drug effects than do prevalence rates, and comparative cumulative incidence tables drug versus active control(s) versus placebo, are very helpful to practitioners.

1. Intrinsic trial design considerations

   An attempt should be made to characterize the patient population susceptible to adverse drug effects. Some extraneous factors can complicate the safety data, such as variations in soliciting and reporting adverse events among the investigators, and differences in the definition of normal ranges for lab values among different laboratories. Since adjustment for their effects may be difficult, precautions should be taken in the design stage of the trial to minimize the influence of these factors by preparing clear and specific instructions for data collection, and monitoring adherence of the investigators and the laboratories to the protocol. Procedures for normalizing laboratory data, for example, may be employed. As previously mentioned, developing standardized toxicity grading scales that may be employed in all studies may also be useful.

2. Adequate numbers

   The ability to detect adverse experiences is dependent on the number of patients evaluated in the clinical trials and in clinical usage. Studies of less than 300 patients per group do not have the statistical ability to necessarily detect adverse experiences in that group of less than 1%. In most cases however, it is permissible to combine studies of equal duration to establish adverse experience rates.

   For any chronically administered product, the safety data base should include at least 300 patients treated with the maximally recommended dose for at least 6 months and at least 100 patients treated for at least 12 months (ICH Guideline for Industry: The Extent of Population Exposure Required to Assess Clinical Safety for Drugs Intended for Long-term Treatment of Non-Life-Threatening Conditions, March 1995 (ICH Safety Guideline)).

   [Issue for Advisory Committee members: What is the appropriate size of the safety database. The CDER Guideline for the Clinical Evaluation of Anti-inflammatory and Anti-rheumatic Drugs (1988) calls for 200-400 patients for one year and 100-200 for two years. (This is considered desirable for the safety evaluation of NSAIDs in particular, because of their known adverse event profile). The

"DMARD" portion of this guideline calls for 400 patients for one year and 200 patients for two years. The ICH Safety Guideline allows exceptions for classes or examples of drugs with known or potential safety problems. To what extent is the ICH recommended safety database adequate for evaluating the safety profile of various RA treatments?]

G. Informed Consent

In each case it is important that informed consent be complete and that patients be able to understand what they are consenting to. If the potential exists for disease exacerbation, this should be part of the informed consent.

III. SPECIAL CONSIDERATIONS FOR BIOLOGICAL PRODUCTS

Although there are similarities between RA trial designs for drugs and biologics, biologics have special characteristics and problems that should be considered in their development.

A. Species Specificity

The schemes used traditionally in determining the initial human dose may not pertain to biologics. Biologic agents may behave differently in animal models than in humans, depending on the physiologic relevance and avidity for the receptor of the ligand in the animal compared to the human.

B. Dose Responses

The dose response curve may be steep (narrow therapeutic window) and/or even hyperbolic, and an agent can be quite toxic at levels just above those thought to show efficacy.

C. Toxicity Response

The toxicity response curve may be highly unpredictable and potentially very dangerous, and include the risk of disease worsening. Biologics may have the potential for disruption of immunologic and physiologic processes. Monoclonal antibodies to cellular epitopes of the immune system, for example, or to TNF receptors, can or may cause serious morbidity at doses only slightly higher than those that are efficacious with markedly less toxicity.

D. Product Homogeneity

This often plays a critical role in activity and toxicity of a compound. Product alterations can greatly affect physiologic activity. Thus, biologics should have consistent lot-release criteria and be reasonably well characterized to be properly evaluated.

E. The Role of Neutralizing Antibodies

If Phase 2 data suggest that agent-induced neutralizing antibodies may interfere with the efficacy of a biologic agent over time, it may become necessary to formally investigate the possibility in a randomized controlled setting. The occurrence of neutralizing antibodies may require reconsideration of doses and dose regimens.

IV. SPECIAL CONSIDERATIONS FOR MEDICAL DEVICES

A. Background

Medical devices for the treatment of RA vary considerably in their therapeutic intent, ranging from agents designed for primary therapeutic effectiveness to those utilized as therapies adjunctive to drugs or biological agents. The variability in therapeutic effects due to disease and response heterogeneity may be more problematic with devices than with drugs and biologics. Preclinical testing requirements cannot be generalized because devices for RA have a diverse range of chemical, mechanical, and electrical properties. In addition, the issues of the optimal placebo control and of local versus systemic effects are common in the evaluation of medical devices. These factors are relevant to both efficacy and safety determinations as described below.

B. Efficacy Considerations

1. Some medical devices intended for local administration may have unexpected systemic therapeutic effects, so precise determinations of mechanisms of action should be made to minimize this phenomenon.

2. Use of a "sham" device is the most desirable placebo control for medical devices, but the success of patient and/or physician blinding with sham devices is not always complete. Blinding may not be feasible if the product is delivered in a surgical or invasive medical procedure. Since inadequate blinding usually biases efficacy determinations in favor of therapy, *design of adequate blinding and its monitoring is imperative.*

3. For devices intended to be utilized as adjunctive therapies to drugs or biologics, design approaches and analysis methods should balance or

account for the differences in disease status and severity, in order to minimize biases in endpoint outcomes. Similarly, the primary therapy with drug or biological agent should be consistent to avoid outcome bias, as should additional, possibly confounding co-therapy (hot/cold therapy, splinting, physical therapy, orthotics, etc.)

4. The issue of quality of life (QOL) determinations is very important for devices intended for rehabilitative purposes, particularly if there are substantial technical demands of certain device uses. Device QOL benefits should be judged by their ease and convenience of administration by assessing the satisfaction with therapy and the improvement in QOL. The outcomes of these determinations should be blinded from the participating investigators to avoid assessment bias.

5. For devices necessitating in-hospital or in-office use, it is recommended that clinical utility be determined accurately and early in development. In addition to adverse event risks, the practical "risks" of the product, such as inconvenience or pain with administration, should also be characterized and judged as efficacy outcomes. Although it is difficult to gather reliable efficacy data, let alone clinical utility, early on, this is critical for the sponsor in order to be able to make a reasoned “go/no go” decision. Agency consultation is advisable.

C. Safety Considerations

1. The availability of well-characterized short-term adverse event rates (3-month cumulative incidence of about 1%), as described for drugs, may not be feasible for medical devices. Due to the more technically demanding administration of devices, it is generally not feasible to enroll large numbers of patients or to conduct several concurrent studies. The timing of device adverse events may differ from that of drugs in that common adverse events may not occur frequently within the first few months of treatment. Therefore, patients with devices which have a delayed effect noted in preclinical or Phase 2 testing should have extended follow-up beyond time on device. These factors may constrain the ability to capture adverse events needed to build an adequate safety database, and may therefore need to be addressed in post-approval studies designed to increase the duration of follow-up or increase the numbers of patient exposures.

2. Because some medical devices are administered in conjunction with a medical or surgical procedure, the distinction between a device-related or procedure-related adverse event is sometimes obscure. The nature, timing, and degree of severity are some factors used to help determine whether an

adverse event is device- or procedure-related. These determinations are often based on clinical judgment, so if blinding is inadequate a potential for bias exists. For this reason, the evaluator should be blinded to treatment (i.e., segregated treating and evaluating physicians). It is recommended that sponsors detail protocol guidelines for assessing procedure-related versus device-related adverse events.

3. Although some medical devices (e.g., those emitting radiation or those administered with a procedure) for RA treatment may be used intermittently, some may be intended for chronic use, so identification of a maximum lifetime exposure or a maximum frequency of exposure to the device is important.

V. SPECIAL CONSIDERATIONS FOR JUVENILE RHEUMATOID ARTHRITIS

A. Background

Juvenile rheumatoid arthritis is a heterogeneous group of diseases which share the common feature of chronic, idiopathic inflammatory synovitis, with onset prior to 16 years of age. These disorders have been divided into clinically distinct subsets based on the extent of joint involvement and extra-articular manifestations: pauci-, poly-, and systemic-onset JRA, as well as oligoarthritis associated with HLA-B27, and they have been further subdivided based on clinical courses.[7] Immunogenetic subsets appear to correlate with these clinical course subsets, and are also distinct from adult RA.[8] (The HLA-B27 subset is not addressed in this document.) Of these various entities, polyarticular JRA is similar in many aspects, particularly in clinical signs and symptoms, to adult RA. While the other JRA subsets are clinically distinct, it is notable that the synovitis seen in any of the JRA subsets appear to be clinically indistinguishable from adult RA, including similar efficacy responses to existing pharmacotherapy (NSAIDs, methotrexate, and prednisone).[9] As only 3-5% of all patients with rheumatoid arthritis develop illness onset during childhood, many investigational therapeutic agents in this population will therefore receive orphan drug status, according to 21 CFR Part 316 - Orphan Drugs. The application of principles in the conduct of clinical trials for adult RA largely applies as well to JRA, and this section only outlines those areas of difference from adult RA.

Conducting drug studies in children is generally necessary and consistent with the expectations of treatment regimens for this disease. Because pediatric subjects constitute a vulnerable population, conducting research involving minimal risk is important. The Committee on Drugs of the American Academy of Pediatrics has published guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations,[10] and general considerations for the clinical evaluation of

drugs in infants and children,[11] both of which should be consulted. Guidelines regarding informed consent and assent of pediatric patients from the Committee on Bioethics of the American Academy of Pediatrics should also be followed.[12] Conducting clinical trials for patients with JRA, and particularly assessing global disease activity and response to therapy, should involve pediatric rheumatologists or adult rheumatologists who have extensive training in pediatric rheumatology and have demonstrated competence in caring for children with rheumatic diseases.

As a general principle children should not be subjected to an agent that has not been first tested for safety in adults. Testing may begin in children, however, when the anticipated benefits based on existing knowledge may justify the anticipated risks. An agent developed specifically for use in JRA (e.g., a biologic agent targeted against a specific pathogenic process which is unique to JRA, and not present in adult RA) may need to be tested first in children, as exposure in adult RA patients or even normal adult volunteers may be unrevealing. If, however, the agent has potential for use in both adult RA and JRA, then, at minimum, pK-pD and initial Phase 1 data (including maximum tolerated dose) should be available for adults prior to the start of testing in children. JRA trials of drugs that are expected to be similar in efficacy to existing drugs, and which do not represent major therapeutic advances or alternative approaches to the basic mechanism of intervention can be delayed until there is extensive efficacy and safety data from either adults or in other pediatric populations.

The need for reliable inferences does not necessitate a placebo control, but randomization and controls should be employed. The choice of control is a function of what is known about the agent at the time and what other treatments are available to potential trial enrollees. If only an active control is used for an equivalence trial, convincing evidence of the efficacy of the active control should be provided, and the test proposed to establish equivalence should be specified. If there have been no prior adult studies, or if the agent under development has a novel mechanism of action or represents an entirely new class of drug, a randomized, double-blind trial, using either a placebo or an active control group of (anticipated) similar efficacy is indicated. Open label extensions to obtain additional data about risk and persistence of benefit are very valuable. The use of active control (standard of care therapy) in the control arm, dose-response design (where control receives a lower dose(s) of the test agent), crossover, or, if the agent has a short onset of effect, randomized placebo-phase trial designs are encouraged as possible alternatives to inactive placebo control in JRA studies. As a general principle, protocol escape clauses are encouraged to permit children who are not responding well to experimental therapy to receive early conventional or alternative treatment. However, when escape clauses are inserted, the sponsor should also indicate how such dropouts will be handled in the analysis.

B. Applicability of the Pediatric Regulation and Impact on Trial Design for JRA Studies.

The "pediatric use" section of labeling regulations (21 CFR 201.57) permits drug and biologic products to be approved for JRA if they have been demonstrated to be safe and effective for adult RA and the disease and mechanism of action of the drug are sufficiently similar in children. Although the regulation allows extrapolation of adult efficacy data, usually additional pediatric dosing and safety evaluations are needed. The following applications of the pediatric labeling rule are applicable to JRA clinical trials. In all cases, application of the pediatric rule may be applied to the signs and symptoms claim only; other claims, including quality of life, radiographic progression, and remission, should have separate JRA efficacy studies. The label should reflect the specific studies performed and documentation provided (efficacy studies in all JRA subsets, or safety and pK studies only in polyarticular JRA, without demonstration of efficacy), in accordance with the regulation.

1. For currently approved agents, including traditional NSAIDs which are cyclooxygenase inhibitors, methotrexate, and corticosteroids, adequate efficacy information exists for all JRA and all JRA subsets. For such agents, a labeling claim could be supported using only pharmacokinetic, pharmacodynamic and safety data in JRA patients, although submission of additional JRA efficacy data is encouraged.

2. For agents currently approved for adult RA, which are not approved for JRA, including auranofin, gold sodium thiomalate, hydroxychloroquine, and pencilliamine, adult efficacy data can be used to support a signs and symptoms claim for polyarticular JRA. There is not adequate data to support extension to all JRA subsets. Pediatric safety and dosing studies of adult data should be submitted to support a label claim for polyarticular JRA. The agency should be consulted to assess the need for any additional studies.

3. For new agents not yet approved for adult RA, adult efficacy data can be used to support a signs and symptoms claim for polyarticular JRA if there is biologic plausibility that the agent would have a similar effect in JRA. When evidence for biologic plausibility does not exist, evidence should be submitted to support the application of the pediatric rule (the agency should be consulted in determining whether adequate biologic plausibility exists to apply the pediatric rule). Pediatric safety and dosing studies should be submitted. The extent of safety testing will depend on the agent, its prior use and any established safety in other pediatric populations. It is

desirable that as much efficacy evidence as possible be gathered during the evaluation of pediatric dosing and safety.

4. It is preferable that efficacy studies be performed in JRA for the signs and symptoms claim, including agents for which biologic plausibility of a similar effect in JRA exists and other categories listed above. Sponsors who seek approval for all JRA should include all JRA subsets in an efficacy study. The data could support a claim for JRA (subsets not specified) provided that the data do not suggest that the agent is ineffective in any one subset. The label should reflect that efficacy was demonstrated, and that the agent is approved for JRA (subsets not specified).

   When the pediatric regulation is applied, the need for pharmacokinetic, pharmacodynamic, and safety studies may still remain. Separate pK-pD studies are not needed for each JRA subset, although all subsets should be represented in such studies. However, due to greater toxicities associated with drug treatment of systemic-onset JRA,[13] [14] [15] strong consideration should be given to conducting studies which allow for stratified analysis of this subset of JRA. If data are available and the coefficients do not differ significantly for adults and children, then the number of time points at which specimen collection is done can be reduced to the minimal number to confirm the curves observed in adults. Micro-sampling techniques should be employed for such studies.

C. Outcome Variables and Claims

It is possible for sponsors to seek approval for all JRA subsets, or to seek approval for individual subsets. In the former case, the label should note the trial numbers in each subset and character of each subset response. Except as noted above in the application of the pediatric rule, all claims should be supported by an efficacy demonstration in the intended subset(s).

1. Clinical Signs and Symptoms:

   All JRA trials should evaluate improvement based on the definition of improvement established by the JRA core set: 3/6 (MD global, parent/patient global, number of active joints, number of joints with limited range of motion, functional ability, and ESR) improved by at least 30% and no more than 1/6 worsening by more than 30%.[16] Protocol individualization may necessitate a refinement in the responder test for patients: for pauci-articular JRA, with, for example, one knee involved and a normal ESR, use of joint and functional assessments specific to the involved joints, and evaluation of uveitis as co-primary endpoints may also

be valuable.[17] For patients with systemic onset JRA, additional assessment of fever, extra-articular manifestations, and thrombocytosis/leucocytosis may be useful co-primary endpoints.[18] Outcome variables need to be clinically "sensible" and appropriate to the type of agent under investigation. Investigators should decide a priori how much change is considered clinically important for each outcome variable.

In all cases, trials should be at least three months, and some assessment weighing all time points equally should be used.

2. Function/Quality of Life

This claim is proposed to reflect demonstrated improvement in function and health related QOL, for six consecutive months, and demonstrated success in signs and symptoms over the same period. This is currently obtainable only in principle, as adequate methodology is not yet at hand. Endpoints will need to be tailored to subtypes enrolled in trials (e.g., to assess knee function in pauci-articular JRA patients who may have this as their primary arthritic manifestation). Instruments should be developmentally validated for the age ranges studied in a trial.[19]

3. Prevention of Structural Damage

Similar to adult RA, this claim would reflect trials of one year or more with concomitant success in signs and symptoms. Currently, only sparse data exist regarding the usefulness of only one radiographic measure in JRA: the carpal-metacarpal distance in those patients with wrist arthritis.[20] Other clinically promising settings include the evaluation of erosive disease in systemics with polyarthritis, hip assessment in systemics, and knee assessments in pauci-articular JRA.

4. Complete Clinical Response

The claim of complete clinical response reflects achievement of six consecutive months of morning stiffness of less than 15 minutes duration, no active synovitis (pain, redness, tenderness to palpation, swelling, stable or decreasing limitation of motion), no extra articular features (including fever, serositis, adenopathy, hepatosplenomegaly, rash, uveitis), and normal laboratory parameters (including ESR, platelets, WBC) and where applicable, no ongoing structural damage while continuing on therapy. Trials should be of one year duration. Residual damage from prior disease, including extra articular manifestations, is acceptable in meeting criteria for complete clinical response. Because complete clinical response rates may

be relatively high in JRA, these studies should be controlled. The need for ongoing therapy may be undesirable if the toxicity of the agent is unacceptable.

5. Remission

Remission is characterized exactly as above, but off drug.

6. Major Clinical Response

[Need Advisory Committee input on this claim and its feasibility]

Patients with chronic synovial thickening without clinically active synovitis (stable synovial thickening) show limited but stable range of motion but may have pain so they would not qualify as a complete clinical response/remission. A "major clinical response" claim for these patients (analogous to this claim in adult RA), represents a response more important than signs and symptoms but less than a complete clinical response/remission. This claim has not yet been fully defined, but it is expected to be a "data driven" definition, similar to the adult RA definition.

D. Trial Design Issues

Recommendations for efficacy studies are based upon the nature of the agent under development. The principles outlined for adult RA are generally applicable. Patients enrolled into these trials may be of any onset or disease course subset. Separate trials for each JRA subset are recommended if the agent is predicted to have a target mechanism of action that will not be applicable and equally efficacious in all JRA subsets. Alternatively, a single, sufficiently large trial with enrollment appropriately stratified provides for useful conclusions to be reached about efficacy and safety for each subset. Co-variates (for adjustment in the analysis) should include, at a minimum, disease course type, disease duration, and non-response to prior methotrexate. Given that JRA is an orphan disease, there is often some flexibility in trial design, but this should be discussed on a case-by-case basis.

At this time, JRA patients should not usually be eligible for entry into efficacy trials unless they have failed to respond adequately to at least one standard "second line agent" (such as methotrexate at a dose of at least 10 mg per meter squared body surface area per week). There may be exceptions to this if, for example, there is evidence that greater efficacy could be obtained by using the agent very early in the disease course, evidence that delayed use in sicker patients potentially carries greater risk of toxicity, or evidence that the agent has a favorable safety and

efficacy profile in a comparable population studied to date and that the agent's actions are potentially readily reversible.

Whether or not the patient continues to receive the agent upon discontinuation from protocol, the patient should be monitored periodically for an extended period. Effects on skeletal growth, development, behavior, sexual maturation, reproductive capacity, and secondary malignancy should be included in the monitoring.

E. Concurrent Antirheumatic Agent Administration

The general principles outlined are applicable in that the goal is to limit the use of discretionary concurrent antirheumatic therapies as much as reasonably possible such that total interpretation of efficacy and safety data is not irrevocably compromised. However, limitations of concurrent medication cannot violate ethically justified treatments nor should it make the protocol so unattractive to parents, physicians, and patients that enrollment is threatened. If background treatment is necessary, early tolerance studies, to ensure safety of co-administration, should precede any large trials.

If patients receive concurrent slow acting or prednisone therapy, the dose should be stable prior to study entry, and preferably remain so throughout the trial. Concurrent medications are usually important prognostically and so may need stratification. If possible, intra-articular steroid injections should be disallowed for a minimum of one month prior to beginning experimental therapy; otherwise that joint should be discounted in assessing therapeutic effects.

F. Multi-centered Trials and Center Effects

Although JRA is the most common rheumatic disease of childhood, its prevalence is low compared to adult RA. Thus, trials of JRA that require large numbers of patients will likely be multi-centered. Multi-centered studies should employ a standardized protocol and data collection forms among all centers. Pretrial meetings of all investigators and other involved personnel are strongly encouraged to assure uniformity in protocol interpretation, patient evaluation, and data recording. Studies have shown that, within a cooperative group, a center's performance is a function of the number of patients enrolled at the center.[21] Thus, studies that use fewer centers with greater numbers of patients at each center are preferable to those that use large numbers of centers with fewer patients. Effort should be made to enroll at least 10 to 12 patients at each center to provide for greater quality assurance. In all multi center trials, center effects should be examined. In such trials, a therapy should show effect in more than one center. When stringent entrance criteria restrict the number of patients eligible for study,

many centers may be unable to enroll even 10 patients. In such situations, randomization blocked within individual centers, rather than across all centers, may help to reduce the potential impact of center effects.

References

1. Felson, D.T., et al. 1993. "The American College of Rheumatology Preliminary Core Set of Disease Activity Measures for Rheumatoid Arthritis Clinical Trials." *Arthritis and Rheumatism, 36 (6): 729-740.*

2. Felson, D.T., et al. 1995. "American College of Rheumatology Preliminary Definition of Improvement in Rheumatoid Arthritis." *Arthritis and Rheumatism*, 38 (40): 1-9.

3. Pinals, R.S., et al. 1981. "Preliminary Criteria for Clinical Remission in Rheumatoid Arthritis." *Athritis and Rheumatism*, 24 (10): 1308-1315

4. Peck, C., et al. 1992. "Opportunities for Integration of Pharmacokinetics, Pharmacodynmics, and Toxicokinetics in Rational Drug Development." *Pharmacological Research*, 9 (6): 826-833.

5. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. *General Considerations for the Clinical Evaluation of Drugs*. 77-3040.

6. U. S. Food and Drug Administration. 1993. *Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs.* 58 FR 39406.

7. Cassidy, J. T., J. E. Levinson, J. C. Bass, et al. 1986. "A Study of Classification Criteria for a Diagnosis of Juvenile Rheumatoid Arthritis." *Arthritis and Rheumatism*, 29:274-281.

8. Nepom, B. 1991. "The Immunogenetics of Juvenile Rheumatoid Arthritis." *Rheumatic Disease Clinics of North America*, 17:825-842.

9. Grom, A. A. and D. N. Glass. 19__. *Juvenile Rheumatoid Arthritis*.

10. American Academyof Pediatrics, Committee on Drugs. 1995. "Guidelines for the Ethical Conduct of Studies to Evaluate Drugs in Pediatric Populations." Pediatrics, 95:286-294.

11. American Academy of Pediatrics, Committee on Drugs. 1982. *General Considerations for the Clinical Evaluation of Drugs in Infants and Children*. Report to the U.S. Food and Drug Administration. Available through the Freedom of Information Office of the FDA. FDA Contract No. 223-79-3003 & 223-82-3009.

12. American Academy of Pediatrics, Committee on Bioethics. 1995. "Informed Consent, Parental Permission, and Assent in Pediatric Practice." *Pediatrics,* 95:314-317.

13. Silverman, E. D., J. J. Miller, B. B. Bernstein. and T. Shafai. 1982. "Consumption Coagulopathy Associated with Systemic Juvenile Rheumatoid Arthritis." *The Journal of Pediatrics*, 103:872-876.

14. Hadchouel, M. A., M. Prieur, and C. Griscelli. 1985. "Acute Hemorrhagic, Hepatic, and Neurologic Manifestations in Juvenile Rheumatoid Arthritis: Possible Relationship to Drugs or Infection." *The Journal of Pediatrics*, 106:561-566.

15. Kornreich, H., N. N. Malouf, and V. Hanson. 1971. "Acute Hepatic Dysfunction in Juvenile Rheumatoid Arthritis." The Journal of Pediatrics, 79:27-35.

16. Giannini, E. H., N. Ruperto, A. Ravelli, D. J. Lovell, and A. Martini. 1996. "Preliminary Definition of Improvement in Juvenile Arthritis." Submitted to *Arthritis and Rheumatism* for the Pediatric Rheumatology International Trials Group.

17. Lindsley, C: Juvenile Rheumatoid Arthritis Workshop, 19__, Proceedings.

18. Silverman, E., et al. 19__. "IVIG in Systemic Onset JRA." *The Journal of Rheumatology.*

19. Murray, K. J. and M. H. Passo. 1995. "Functional Measures in Children with Rheumatic Diseases." *Pediatric Clinics of North America*, 42:1127-1154.

20. Hare, L., L. Wagner-Weider, A. Poznanski, et al. 1993. "Effects of Methotrexate on Radiologic Progression in Juvenile Rheumatoid Arthritis." *Arthritis and Rheumatism*, 36:1370-1374.

21. Sylvester, R. J., H. M. Pinedo, M. De Pauw, et al. 1981. "Quality of Institutional Participation in Multi Center Clinical Trials." *New England Journal of Medicine*, 305:852-855.